DOCUMENT RESUME

ED 238 934 TM 840 036

AUTHOR Burry, James
TITLE Federal Requirements Affecting Local Evaluation Use. Evaluation Productivity Project.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Dec 83
GRANT NIE-G-82-0001
NOTE 37p.
PUB TYPE Reports - Research/Technical (143) -- Viewpoints (120)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; Evaluation; *Evaluation Criteria; *Evaluation Methods; *Evaluation Utilization; Evaluators; *Federal Regulation; Performance Factors; Researchers; School Districts
IDENTIFIERS *Use Factors

ABSTRACT

Use factors comprise those characteristics or elements present in evaluation settings that potentially affect use. This paper focuses on one such use factor, termed "written requirements for the evaluation." This factor includes the evaluation requirements accompanying federally-funded programs. Two questions are dealt with: how does this factor interact with other factors in evaluation, and is this interaction likely to have a positive or negative effect on use? What research has uncovered about the factors affecting local evaluation use is discussed, and the federal program evaluation requirements currently in effect are described. An ideal scenario is presented in which the factors have a high potential for stimulating local evaluation use. Finally, the impact of federal requirements or factors promoting evaluation use are discussed. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED238934

DELIVERABLE -- DECEMBER 1983
EVALUATION PRODUCTIVITY PROJECT

Federal Requirements Affecting Local Evaluation Use

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. Gray

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Marvin C. Alkin
Project Director

Grant Number
NIE-G-82-0001, P2

CENTER FOR THE STUDY OF EVALUTION
Graduate School of Education
University of California, Los Angeles

TM 840 036

FEDERAL REQUIREMENTS AFFECTING LOCAL EVALUATION USE

Report Prepared by:
James Burry

Center for the Study of Evaluation
University of California, Los Angeles

December 1983
Regents of the University of California

Table of Contents

Factors Affecting Local Evaluation Use	2
Federal Program Evaluation Requirements	3
Chapter 1 Evaluation Requirements	5
Chapter 2 Evaluation Requirements	5
Bilingual Program Evaluation Requirements	6
Education for the Handicapped Evaluation Requirements	7
Vocational Education Evaluation Requirements	8
Broad Problem Areas	9
An Ideal Scenario	11
How the Scenario Factors May Promote Local Evaluation Use	15
The Effects of Current Federal Requirements	18
Chapter 1 Evaluations	18
Chapter 2 Evaluations	20
Bilingual Program Evaluation	21
Handicapped and Vocational Education Evaluations	23
Some Observations from the Field	24
Summary	27
Figure 1: Factors Affecting Evaluation Use	30
References	33

FEDERAL REQUIREMENTS AFFECTING LOCAL EVALUATION USE

> . . . As the shift continues from the federal to the state levels in the management of education programs, the states become more, not less, accountable for them. SEAs and LEAs have become accustomed to the federal government not only requiring the evaluation of programs but also dictating methods of evaluation (Council of Chief State School Officers, 1982).

For the past four years, the Center for the Study of Evaluation (CSE) has focused on identifying the factors which affect local uses of evaluation. As defined here, the term use refers to the planned actions that result from applying evaluation information and processes to the resolution of specified problems, questions, or concerns. While evaluation can have unplanned impacts and consequences, our research has emphasized only those which are intended to take place. The analysis in this paper is further limited to use that occurs at the local program level.

Use factors comprise those characteristics or elements present in evaluation settings that potentially affect use. Our research has identified a number of such factors, and this paper focuses on one of them. Termed "written requirements for the evaluation," this factor includes the evaluation requirements accompanying federally funded programs. Two questions concern us here: how does this factor interact with other factors in the evaluation? and is this interaction likely to have a positive or negative effect on use?

After discussing what our research has uncovered about the factors affecting local evaluation use, we will describe the federal program evaluation requirements currently in effect. Drawing on our research findings,

we will then present an ideal scenario in which the factors are such that they have a high potential for stimulating local evaluation use. Finally, we will discuss the impact of federal requirements on factors promoting local evaluation use.

Factors Affecting Local Evaluation Use

Our research on evaluation use falls into four phases. The first phase, case studies of school-level program implementation and evaluation, identified some of the factors affecting use and led to an agenda for further research (Alkin, Daillak, & White, 1979). The second phase, field studies of evaluators at work, elaborated some of these factors, especially those reflecting the organizational context of the program being evaluated (Daillak, 1980). The third phase involved a user survey which elicited a picture of the kinds of decisions educational administrators make and the kinds of information they use as these decisions are shaped (Stecher, Alkin, & Flesher, 1981). In the fourth phase, we synthesized what we know about evaluation use, about the range of factors affecting use, and about the ways in which these factors can be influenced to promote use (Alkin et al., 1983). This paper draws on that synthesis.

Figure 1 classifies and lists the factors that potentially affect the local uses of evaluation. The individual factors are grouped into three broad categories. Human factors reflect the evaluator and user characteristics affecting use: their attitudes toward and interest in the project and its evaluation, their backgrounds and organizational positions, and their professional styles. Context factors are inherent in the setting of the particular program being evaluated; they include the characteristics of that program, the requirements and fiscal constraints imposed on the

evaluation, and the interrrelationship between the program and other units of the larger organization and the surrounding community. Evaluation factors relate to the actual conduct of the evaluation: the evaluator-user interaction, the procedures used, and both the substance and the format of the information reported to potential users. (Each of the factors listed in Figure 1 is defined and exemplified in Alkin et al., 1983.)

All these factors can affect evaluation use. For instance, the role chosen by the evaluator and his or her ability to generate locally relevant information help to determine whether the findings of the evaluation are ever applied in decisions about the program. Moreover, one factor can influence--and can be influenced by--other factors in the situation. For instance, the evaluator's ability to focus on local needs may be constrained by other factors in the setting. It may be that federal requirements cause the evaluator to adopt a certain role and to collect certain kinds of information and thus make it difficult for him or her to focus on local information needs.

Federal Program Evaluation Requirements

Here we examine the requirements accompanying three major types of federal programs: programs existing under Chapters I and II of the Education Consolidation and Improvement Act of 1981, and those few programs still existing as separate entities.

The Education Consolidation and Improvement Act (ECIA) incorporated most previously existing categorical programs into a few large block grants. Title I of the Elementary and Secondary Education Act (ESEA) of 1965, authorizing education for disadvantaged students, became Chapter 1 of the ECIA.

Chapter 1 retains the intention of the Title I legislation which it replaced. Under Chapter 1, a state receives block funding for the purpose of providing services to meet the needs of educationally deprived children. The formula and eligibility criteria established under Title I remain in effect. A local school district may receive a grant under Chapter 1 if it files an application with its state education agency and if the agency approves the application. The district must use the grant only for the purposes set out in Chapter 1.

Chapter 2 of the ECIA consolidated most of the programs authorized under Titles II, III, IV, V, VI, VIII, and IX of the ESEA; parts of the Title V of the Higher Education Act of 1965; and Follow Through (on a phased basis).

The block funding received by a state under Chapter 2 is based on the ratio of its school-aged population to the total school-aged population nationwide. Chapter 2 authorizes a wide variety of activities, including basic skills improvement, metric education, health education, school library services, community schools, safe schools, and ethnic heritage, to name but a few. A local school district receiving funds under Chapter 2 may elect to use its grant for any one or some combination of these authorized activities.

A few of the original categorical programs remain as separate, non-consolidated entities: PL 95-561, Bilingual Education; PL 94-142, Education for All Handicapped Children; PL94-482, the Vocational Education Act. For programs such as these a local school district may apply for and, if its proposal is successful, receive federal funds.

The three kinds of programs discussed here--Chapter 1, Chapter 2, and the nonconsolidated programs--differ substantially in their federal evaluation requirements. Chapters 1 and 2 are intended to produce evaluations which will eliminate burdensome and unproductive paperwork and free local schools and districts from unnecessary federal supervision and control. The federal requirements accompanying the nonconsolidated programs remain in the form they had before passage of the ECIA.

Chapter 1 Evaluation Requirements

Chapter 1 requirements specify that programs be based on an annual needs assessment of educationally deprived students. The evaluations of such programs will use objective measures of basic skills achievement to determine the extent to which student performance is improved under the program. Local schools will keep the records necessary for fiscal audit and program evaluation.

At the local school level, the requirements for evaluation are as follows:

- Programs will be evaluated in terms of their effectiveness in achieving specified goals
- Evaluations will be based on objective measures with the matter of test selection/development to be decided on by the local educational agency
- Sustained effect evaluations should be conducted to determine whether improved performance is maintained over a period of more than one year

Chapter 2 Evaluation Requirements

No firm evaluation requirements are prescribed under Chapter 2. State and local applications for funds must, however, include agreements to keep records and to provide information as may reasonably be required for fiscal

audit and program evaluation. (Beginning in fiscal 1984, applications must provide for an annual evaluation of program effectiveness, the results of which will be made available to the public.)

Requirements under Chapter 2, then, are much less stringent than under Chapter 1. They are also a far cry from the detailed evaluation requirements connected with some of the original categorical programs now subsumed under Chapter 2.

Bilingual Program Evaluation Requirements

Bilingual program evaluation requirements are more prescriptive than those of Chapters 1 and 2. Proposals for funds must be accompanied by evaluation plans containing the following components:

- Provisions for measuring the accomplishment of the project's instructional objectives
- Provisions for measuring students' progress in improving their English-language skills
- Procedures for using the information gained from the evaluation to improve the project's operation

Further, districts submitting proposals are likely to place varying emphases on the federal criteria used to review proposals. These criteria consist of the following components amplifying the basic elements of the required evaluation:

- The overall evaluation design should be appropriate for the stated objectives of the project
- The design should give attention to the assessment of the extent to which all of the project's objectives have been achieved, to data collection methods and instruments, to data analysis procedures, to time schedules, and to staff responsibilities
- The evaluation design should specify a comparison procedure for estimating what the performance of the participants would have been in the absence of the project
- Methods must be specified for identifying nonparticipants for comparison or for utilizing some other comparison standard, such as historical or statistical

- Sampling procedures must be identified, if applicable, and steps must be taken to ensure that sample results will be representative of the entire project population
- Data collection and analysis procedures which address the evaluation questions and which are appropriate for use with the project data must be identified

Education for the Handicapped Evaluation Requirements

Programs for the handicapped need comply with a smaller number of requirements than bilingual programs. Moreover, their evaluation requirements seem to differ from those of bilingual programs with respect both to specificity and to apparent end result.

As stated in the federal requirements governing this program, the purpose of the evaluation is to assess the effectiveness of efforts to educate handicapped children. States and local districts have differing responsibilities in this regard.

States are responsible for the general supervision of all special education programs and their evaluations. Specifically, states are required to monitor how handicapped children are identified, located, and evaluated at the local level. ("Evaluation" here refers to the diagnostic assessment of each student for the purpose of designing an Individual Education Plan--IEP--for that student.)

Local school districts are responsible for the actual work associated with the IEP's, each of which must indicate:

- The student's present level of performance
- Annual goals for the student, including short-term measurable objectives related to long-term goals
- Specific educational services to be provided for the student, and the extent to which the student is able to participate in local educational programs
- Appropriate and objective criteria and evaluation procedures and schedules to be used in the annual review to determine whether short-term objectives have been achieved

Tests and other procedures used to assess students locally must be validated and administered by trained personnel. These measures must assess specific areas of educational need, rather than providing some broad index of intelligence. No single measure is to serve as the sole criterion for assessing students.

States must evaluate programs at least once a year to determine how effective they are in meeting the educational needs of handicapped students. The IEP's must also be evaluated. As part of their monitoring responsibilities, states must collect data and school reports, conduct on-site visits, and compare a sample of the IEP's with the programs actually provided. The state must also prepare an annual report indicating the number of students served, by disability group.

Vocational Education Evaluation Requirements

The requirements associated with vocational education programs are less rigorous than those associated with bilingual programs and programs for the handicapped. Indeed, school districts receiving federal funds for vocational education are subject to only one federal evaluation requirement: they must evaluate each program every five years, following directions and procedures specified by the state education agency. Typically, a certain percentage of secondary schools is sampled each year, so that each school is surveyed once over a five-year period.

These evaluations often take the form of a student follow-up survey. That is, graduates of vocational programs are sent a questionnaire asking them to provide information on such questions as whether they are working on a job related or unrelated to their training, what their entry-

level earnings are, and whether they are satisfied with the training they received. In addition, employers may be asked if they are satisfied with the entry-level skills of those students working in a job related to their training.

Broad Problem Areas

Having sketched the kinds of federal evaluation requirements associated with Chapters 1 and 2 programs and with the major programs still operating outside the ECIA, we turn now to a consideration of some broad problem areas connected with the language of the requirements.

The first three problem areas we are concerned with, although they are associated with Chapters 1 and 2 federal requirements, do not stem directly from the language of these requirements; they are not problems of commission. The three areas reflect the antecedents to the current federal evaluation regulations and issues stemming from increased local responsibility and decreased resources.

The first two problem areas are based in antecedents to current requirements. First, some potential evaluation users may not be convinced that recent changes in requirements will actually reduce external supervision and control. That is, while federal supervision may decrease, other agencies intervening between school districts and the federal level, such as the state, may continue or initiate policies which offset the freedom that is intended to result from the new federal requirements.

Second, some potential evaluation users may be so thoroughly immersed in the business of administering a previously required test that they remain unaware of their new freedom. Others may be uncomfortable with the additional responsibilities accompanying relaxed requirements. Even if

other agencies intervening between the school district and the federal level try to encourage new approaches in evaluation, the strength of previous practice may thwart these efforts.

The third problem area reflects increasing evaluation freedom and responsibility at the local level coupled with decreasing resources. On the one hand, with fewer regulations to follow there is increased demand for local decision-making responsibility for planning and conducting evaluations. On the other hand, with reductions in resources, local school districts may be unable to supply sufficient evaluation expertise in all program areas (as was recently pointed out by Gray, Caulley, & Smith, 1982). A local evaluator may need to take responsibility for simultaneous evaluations of Chapter 1 and 2 programs and of other programs such as bilingual education.

While the above problem areas cannot be ascribed directly to the language of current federal evaluation requirements, others can. For example, some regulations, such as those accompanying bilingual programs, require or are locally understood to require so much for external reporting purposes that little room is left for maneuvering to meet local information needs.

Related to the above issues is the emphasis on specific technical procedures in the regulations accompanying some programs. This potential problem area, though it is of particular concern in bilingual programs, is seen to some extent in the language of the Chapter 1 regulations. Coupled with dwindling local resources and evaluation expertise thinly stretched, a school district facing multiple program evaluation needs of a technical nature may find it difficult to comply.

An Ideal Scenario

With these general problems to serve as a background, we will now sketch an "ideal" scenario: a picture of an evaluation situation in which the factors are such as to increase the likelihood of local evaluation use. Since this scenario will be used later in discussing influences on use, however, it is tempered by reality. That is, it includes factors which can constrain use as well as factors that can promote use (though weighted more heavily in favor of the latter).

> A school district received a grant from the Bureau of Bilingual Affairs* (BBA) to design and pilot test a bilingual education program for students recently arrived in the United States from a country in Central America. The instructional program supported by the grant was to stress English-language acquisition and rapid student transition to English-language instruction in the basic skills. The budget set aside funds for a full-time program evaluator to be committed solely to the new program.
>
> The grant covered thirty months. The first six months were devoted to planning and designing the program, which was then implemented for two years in a pilot school serving substantial numbers of the target population.
>
> The BBA required that end-of-year evaluation reports be submitted to its program monitor. Criteria for program success were to be derived from a few broad indicators of basic skills achievement set by BBA. From these broad indicators, the district would develop student performance objectives and then select or develop measures of these objectives. Selection of program evaluator and choice of measures were left to district discretion, with the measures subject to final approval by the BBA.
>
> The first-year evaluation was to be used primarily by the district for making any necessary program revisions before the second-year implementation. The second-year evaluation was to be used by the BBA in deciding if the program should be refunded for further implementation.

---

* The Bureau of Bilingual Affairs is a fictitious agency. It is used here as a means of establishing relationships between a program and a funding agency, and the evaluation regulations which are part of these relationships.

The district chose the assistant director of special programs to be the BBA project director. This administrator has a strong background in bilingual education, takes an active role in local evaluations of the district's special programs, and makes extensive use of evaluation information in carrying out her responsibilities. Further, she has a reputation for being outspoken, being committed to educational excellence, and taking a "can do" approach.

Since the BBA placed no restriction on the choice of evaluator, the project director, after talking with teachers in the new program, designated a member of the district's research and evaluation staff as project evaluator. This evaluator once taught in the district's traditional bilingual program, is well-known and respected by most bilingual teachers in the district, and is frequently called on, formally and informally, to help them with instructional and test-related matters.

Although the evaluator completed a doctorate in research methods after joining the district office, he is fully aware that test information is only one of the many kinds of information which teachers use in their classroom work.

The project director and the evaluator met several times to discuss who should be involved in the BBA project evaluation. They tried to anticipate the questions and needs that teachers (including those not in the program), parents, curriculum developers, central office administrators, and the BBA project monitor might have. They also considered how these questions might best be stated and answered and what kind of information would be most acceptable and useful to the various groups involved.

The project director and evaluator also considered the role of the bilingual office of their own state department of education. At a meeting with the department's director of bilingual education, the project director and evaluator discussed the features of their program and its evaluation. The outcome of this meeting was that, while the district must continue to comply with state regulations governing the provision of bilingual education, state reporting requirements were waived for the duration of the program. Of interest to the state was the question of whether the program's evaluation might offer features which could be considered for statewide adoption.

After these meetings, the project director and evaluator decided that, in addition to the project director, the people who should be most heavily involved in the evaluation were those who could benefit most directly from its findings: the BBA monitor, district curriculum developers, and teachers in the new program. They also decided to provide summary information to other groups not so directly concerned: teachers in the traditional bilingual program, central office administrators working in bilingual education, and parents.

At the beginning of the six-month planning stage, the project director and evaluator set up a series of meetings with district curriculum developers, teachers in the new program, representative teachers from the traditional program, and district officials working in bilingual education.

During these meetings, the project director and evaluator explained the new program and its evaluation needs. The group's first task was to develop a program plan and objectives. Next, with advice from the evaluator and project director, a set of evaluation questions, representing various interests and viewpoints, was agreed upon.

The evaluator asked program teachers and others what kinds of information they needed and when they needed it to answer their questions about the program. Teachers' routine assessment practices were built into the evaluation plan.

Having comparative freedom in their choice of measures, the project director and evaluator decided to apply a test already in use in the district as one means of determining student language proficiency. Teachers were free to adjust the test-based proficiency classifications if adjustments seemed warranted on the basis of students' classroom performance. Teachers agreed to document any such adjustments to help the evaluator with program description.

Since most of the planning participants agreed that a previously required norm-referenced test served no immediate instructional purposes, they decided to develop several formative measures of progress on specific objectives and an end-of-year measure assessing these objectives more broadly. The evaluator took great care to convince staff that their use of non-traditional measures would be acceptable to the funding agency.

Draft measures, along with a description of the objectives each was intended to address, were sent to the BBA monitor, who had them reviewed by a bilingual specialist and a measurement expert.

The final evaluation plan submitted to the BBA (with a copy to their state's bilingual education director) included procedures for administering measures and sending results to the evaluator, who would then interpret the information and report his findings and recommendations. Curriculum developers would consider how the information might be used to modify the program's instructional materials. Teachers would consider necessary changes in classroom practice. Both groups would document these changes for possible inclusion in the end-of-year report.

During the implementation phase, the evaluator and the project director spent a great deal of time working with staff. In addition to attending formal meetings to discuss progress, the evaluator made himself available on an informal basis to answer staff questions, explain recommendations, and suggest how they might be applied.

During the implementation phase, the evaluator and the project director also conducted meetings with the parents of children in the new program, as well as with representatives of the parent group from the traditional bilingual program. While these meetings helped inform both groups about program success, they were intended primarily to minimize any problems, particularly from traditional program parents, about the language approach chosen for the new program. Both the evaluator and the project director were concerned that, if traditional program parents were misinformed about the purpose of the new program, they might try to get the new program parents to support them in any possible disagreement over language transition versus language maintenance.

Whether reporting informally or informally, the evaluator presented his findings and ideas in a variety of ways. He relied not only on his own expertise, but also on the knowledge he had gained about individual staff needs and preferences. During all of the formal sessions, and many of the informal ones, the project director helped resolve issues to ensure that the overall intentions of the evaluation were kept in perspective and that the evaluation proceeded on schedule.

All reports were reviewed by the various groups previously identified. Each individual or group was encouraged to criticize them and to offer feedback which might, at the discretion of the evaluator, be incorporated into the final report.

How the Scenario Factors May Promote Local Evaluation Use

The requirements governing the factors described in this ideal scenario have a feature or two which may not seem too relevant at the local level: that student performance criteria be developed from BBA-established indicators of basic skills achievement; that the measures selected or developed be subject to BBA review.

On the other hand, the requirements do reflect concern with local control of the direction and form which the evaluation might take. In this regard, the BBA stressed that first-year evaluation findings focus primarily on program refinement. This focus, further, helps the local program define its range of measurement needs.

Related to the above is the role of the state department of education in the scenario setting. Importantly, the freedom permitted by the funding agency regulations will not be hampered by state requirements offsetting that freedom or running counter to the evaluation focus and procedures chosen by the program.

Further, the evaluator paid careful attention to the historical dominance of certain kinds of tests as he worked with program staff to convince them of the desirability as well as the feasibility of developing locally-relevant measures of program progress and success.

The need for increased local evaluation responsibility and expertise, a need which is a counterpart to increased local evaluation control, has also been considered. The funding agency set aside budgeted funds for a full-time evaluator who will not have to spend time working on the evaluations of other district programs.

In addition, the funding agency's requirements are not all-consuming, nor can they be interpreted as being all-consuming. They do not mandate, or suggest, so many tasks for external reporting purposes that the program staff will feel they have inadequate time or resources to work local uses into the evaluation. Similarly, though the funding agency is concerned with the technical adequacy of the evaluation, which is probably reflected in their providing funds for a full-time program evaluator, their approach to this matter is quite reasonable. Rather than requiring or suggesting certain kinds of tests, for instance, the BBA elected to let the local program decide what form a test should take in order to be both technically and contextually adequate. Once these tests were developed, the BBA provided a valuable resource by having them reviewed from the standpoint of their content and measurement properties. It appears that technical matters are seen as a means toward an end -- and that end is the successful resolution of the questions guiding the program's evaluation.

Further, the requirements do mention uses. While these uses may have some external application, they also emphasize local program monitoring and so can provide a springboard for consideration of other local uses. The project director and the evaluator must consider how the external uses can be tied in with local uses, and decide upon who the local (and other) users might be. They can then build the evaluation around these users.

Within the general context and direction set by these requirements, there is a great deal of freedom at the local level to ensure that the specific factors important to use in that particular setting receive attention, and there are sufficient time and resources to capitalize on that freedom. Since the requirements neither over- or underspecify but

instead offer an initial focus, the project director and evaluator can devote much of their time to local needs without losing sight of external needs. They can consider and select a set of potential users; find out their interests, questions, and information needs; and build these features into an evaluation that will meet their own local needs. And there is an important spin-off here. By demonstrating her own interest in and commitment to the evaluation, the project director can inspire others to follow her example. Moreover, the evaluator, who enjoyed some credibility before the evaluation, can expand this credibility. Because they do not see him as an external judge, the program staff will be less suspicious of and anxious about the purposes of the evaluation.

The evaluator, with advice from the project director, can spend time analyzing users' needs, interests, and commitment to evaluation. He can also familiarize himself with the political climate, in order to identify any issues which could have a negative effect on use. Through careful planning followed up by meetings with users, the evaluator and project director can design an evaluation which capitalizes on users' interests. The evaluator can then conduct an evaluation whose processes have a high potential for generating the right kind of information at the right time for the right people. For instance, given the time at his disposal, the evaluator can continue to engage in ongoing dialogue with users. This dialogue, aimed at developing a shared perspective on what the evaluation is to accomplish, helps ensure that it will provide users with substantial and relevant information. The project director, given her commitment to and understanding of evaluation in its broader political context, can spend her time ensuring that the evaluation maintains a balanced -- internal-external -- focus.

The kinds of use-promoting activities mentioned above, involving project director, evaluator, and program staff, require certain abilities in and attitudes toward evaluation. It is critical that evaluation requirements do not work against either. In the case discussed here, the program's surrounding context is such that project evaluator and director, both of whom are committed to evaluation use, do not need to fear that external requirements will work against local use. Thus, they maintain their commitment at the same time as they involve other staff in the evaluation, thereby increasing the liklihood that the evaluation will actually be put to use.

The Effects of Current Federal Requirements

Now we will take a closer look at how current federal evaluation requirements compare with those described in the ideal scenario.

Chapter 1 Evaluations

As we pointed out earlier, the federal requirments for Chapter 1 program evaluations involve assessing the achievement of specified goals on the basis of locally chosen objective measures and determining whether improved student performance is sustained over a period of more than one year. These requirements, intentionally or not, may create obstacles to local evaluation use.

For instance, the "objective measures" mentioned in the Chapter 1 requirements may not be uniformly understood. Owing to historical precedent, one local program may believe that a test must be norm-referenced in order to be considered objective and may continue to use this kind of test even when it serves no relevant local purpose. If this is the case, the evaluation findings generated by use of this test are likely to be seen by program staff as having little practical value for them.

In another district -- say, one that is trying to take advantage of the new freedom permitted under current regulations -- the pros and cons of various kinds of tests may be discussed at length without ever resolving the matter to the satisfaction of all potential users. So much time may be spent on the debate over the test question that the evaluator has little time left for planning and conducting an evaluation around locally important issues.

Similarly, attempting to demonstrate "sustained effect," as required by Chapter 1, may create technical problems involving student mobility, design, and sufficiency of data base for districts. As with test-related issues, technical issues should not be allowed to dominate an evaluation. However, they will require substantial amounts of local evaluation expertise, and because that expertise may be thinly spread, it is possible that locally-relevant questions will lose in the resultant competition for evaluator attention.

In the ideal scenario, the matters of testing and technical adequacy were not taken lightly, but neither were they allowed to dominate, to consume valuable resources and energy. The evaluator and the project director were able to devote their time to the task of building uses into the evaluation. They developed processes with an eye to monitoring and improving potentially hostile attitudes toward the program. They legitimized routine assessment practices by placing them on an equal footing with "formal" testing. Thus, they increased the evaluation's credibility and, at the same time, increased the liklihood that the information it produced would be used to answer the programmatic questions about instructional materials and classroom practices of interest to the staff.

In districts meeting the kinds of problems outlined above, that is, where testing issues are difficult to resolve, where there is limited expertise to consider technical adequacy and local relevance, there is likely to be some negative effect on factors promoting use: local anxieties and sense of risk may dominate; program staff may believe their questions and concerns are receiving insufficient attention; the evaluator's credibility is likely to suffer; local commitment to evaluation use and perception of local usefulness will decrease.

Chapter 2 Evaluations

The current requirements for Chapter 2 evaluations permit much greater flexiblity and decentralization than were allowed in the past: they simply instruct state and local agencies to keep whatever records and information that might reasonably be required for fiscal audit and program evaluation. Of course, the juxtaposition of "fiscal audit" (often seen as potentially punitive) and "program evaluation" may itself have a negative effect on local use. But Chapter 2 regulations can affect local evaluation use in other ways.

Some districts may opt to conduct no program evaluation. Others may keep records of whatever tests or other measures were administered but make no effort to analyze and use their results. Still others, accepting the spirit of the requirement, may try to evaluate their Chapter 2 funded activity or activities. In doing so, they face several possible problems.

For example, a district operating a Chapter 2 program that emphasizes only one of the authorized content areas may encounter any of the barriers to uses already mentioned. On the other hand, a district that commits Chapter 2 funds to several of the authorized areas not only faces the

normal problems but may also encounter another dilemma if it tries to evaluate all of its Chapter 2 activities: namely, how to conduct evaluations of several content areas while maintaining a balance among them, and build use potential into each.

For instance, the staff of one activity may feel that it is competing with other staffs for resources, and may fear that it will not get a fair share of attention. Further, if one person is selected to evaluate all the activities in operation, his or her credibility may be suspect; but if each authorized activity has its own evaluator, administrative and management conflicts are likely to arise. These problems, to the extent that they limit efforts to stimulate use-conducive factors, can have a negative effect on the evaluation and its uses.

Other problems may arise because of the varied nature of the activities that can operate under Chapter 2. Some content areas, such as school library services, are candidates for resource expenditure monitoring. Other areas, such as consumer education and ethnic heritage, are more amenable to evaluation. However, consumer education, which primarily emphasizes cognitive achievement, requires a different kind of evaluation than an ethnic heritage program which is likely to emphasize student attitudes.

In short, it seems doubtful that a district will be able to mount simultaneous evaluations of the kind suggested in the scenario without more careful guidance than is presently available under Chapter 2.

Bilingual Program Evaluation

Because the language of current federal regulations governing bilingual education seems to stress the technical aspects of evaluation,

districts often assume they will be penalized if they do not follow classic canons of methodology.

In this context, the regulations use technical terms -- such as "design," "data collection and analysis procedures," "comparison procedures ... to estimate performance," "statistical comparison," "sampling procedures" -- which are intended as suggestions about how a local program might be evaluated. But the people involved in local bilingual programs see them not as suggestions but as federal requirements (or, at the very least, federal preferences) and so spend great amounts of time and other resources trying to accomplish what they feel is expected of them (Burry, 1979).

In our ideal scenario, which centered on a bilingual program, there were few technical "suggestions" about how the evaluation might proceed. Rather, the overall tone of the requirements suggests agency concern with ensuring the liklihood that the evaluation, beyond answering agency questions, will also address the local program's evaluation needs and range of information uses. And these uses are typical of the interests of bilingual program staff -- improving instructional materials, modifying classroom practices, developing a harmonious relationship with the community. Further, the presence of full-time evaluation expertise will help ensure that these program interests receive a balanced treatment, balanced in terms of how the evaluator distributes his time and talent among them, and balanced in terms of the attention they receive in relation to external interests and information needs.

In the ideal scenario, then, we encounter <u>one</u> evaluation designed to meet a variety of needs, one of which is a direct focus on local uses. However, to get relevant information in the appropriate form and at the

right time, many bilingual programs mount parallel evaluations. They conduct a formal evaluation for federal purposes which involves administering (typically) a norm-referenced test, comparing the test results with some (statistical, historical) standard, estimating their program's contribution to student outcomes, and producing a final report. Because program staff, especially teachers, find this formal evaluation of little use for their purposes, their resentment toward it builds as they do their own informal evaluations to get the information they need to run their classrooms.

Handicapped and Vocational Education Evaluations

The federal evaluation requirements associated with these programs are discussed together for two principal reasons. First, handicapped and vocational education programs have traditionally not been subject to evaluation and its uses as defined in this paper. The nature of the populations served, the types of assessment devices required, the kinds of instruction provided, and the outcomes expected are such that attention is directed to the individual student rather than to the program as a whole.

Second, while they differ in the language used and specificity of direction, each program's regulations seem to culminate on the same note: reporting on the numbers of students served and on how well they are served. This note is appropriate: the desired outcome of vocational education is the successful employment of the individual student; the desired outcome of education for the handicapped is the provision of instruction and other services appropriate for the individual child. In both cases, then, the individual -- not the program -- is emphasized.

It does not follow from this, however, that evaluation information cannot be used at the program level. Indeed the regulatory language for both handicapped and vocational education contains the word "program" in association with "evaluation." But as evaluations are currently conceived and conducted, the state, which is the responsible body, seems to emphasize a regulatory function stressing compliance. In the case of education for the handicapped programs, states conduct evaluations which include site visits and which report on the number of students served in each of several disability groups. In the case of vocational education programs, states may require or suggest the use of follow-up procedures such as surveys asking graduates about their current jobs, earnings, satisfaction, and so forth; some states also send site-visit teams to observe classrooms and inspect equipment.

In neither case is there any explicit concern for the kinds of program-level evaluation aggregation illustrated in the scenario. Further, states and local districts seem to have little conception of this kind of evaluation and its uses. Should evaluation as treated in this paper be the intention behind federal requirements, then the requirements will probably have to be reconceptualized.

Some Observations from the Field

To test our impressions of the effects of current federal regulations on local evaluation use -- particularly with respect to the obstacles they create -- we conducted some interviews with the people most responsible for complying with federal regulations: directors of school district research and evaluation units. Most of the people we spoke to agreed that, while current federal regulations do not necessarily preclude use, they offer

little that helps to stimulate local evaluation use; their effect may, at best, be neutral. By implication, these district personnel were saying that, if it is to be a valuable partner in evaluation and its uses, the federal government must do more to promote evaluation use among the local schools and districts whose efforts it is supporting.

While directors agreed that the increased flexibility provided by Chapter 2 and (to a lesser extent) Chapter 1 is desirable, some felt that the new regulations were too nondirective. In this regard, Hastings' (1983) interviews with school district Chapter 2 administrators suggest that the feelings expressed by our respondents are felt elsewhere. Districts need some assistance, our respondents believed, in reconciling what is perceived as required for federal purposes and what is needed for local use. This concern is reflected in the following observations, especially those on historical precedent and on the perceived need to mount parallel evaluations.

Several directors were concerned about the force of historical precedent. Because of earlier state and district preferences (which grew out of earlier federal emphases), they were still required to administer a norm-referenced test for reporting purposes. But information produced by such tests was of limited use to them in carrying out their responsibilities. Further, all but one of the respondents agreed that these tests did not address teachers' interests or provide the kinds of information needed to monitor and adjust classroom instruction. Most directors felt that a good deal of their limited time and resources were given over to generating information which was state required in response to Chapter 1 regulations.

A few districts were trying to distill some locally-useful information from a norm-referenced test. But the directors felt that too much time was involved in having the tests scored and then returned for analysis and interpretation so that they would be of some use to curriculum specialists and classroom teachers. Consequently, these efforts detracted from the time they needed to generate information more specific to local instructional needs.

Several districts, in addition to administering and reporting the results of a norm-referenced test, used locally developed objectives-based measures and/or the tests accompanying curriculum materials. While districts found the information from these additional tests instructionally useful, they asked why they had to conduct what was, in effect, a parallel evaluation. That is, for the reasons suggested above, they felt they should use a "respectable" norm-referenced test for external reporting purposes; given the limitations of the information provided by such tests, however, they felt at the same time that they had to resort to other devices for locally useful information.

The question raised by the directors was as follows: if such district-level efforts are necessary to produce formative data for district purposes, cannot these efforts also be used to satisfy external (summative) requirements? They asked this question even though they were aware that federal requirements imposed no particular kind of test; perhaps they (or their superiors) continue to believe that anything other than a standardized test is unacceptable.

This situation created a double dilemma. On the one hand, the research and evaluation directors realized that, to be locally useful, the evaluation should provide different types of information for different

groups of users and that such an effort takes time: time to identify the needs and questions of various potential uses, time to develop or select appropriate tests or design other data collection procedures, time to win user support for the evaluation. On the other hand, the felt need to run a separate evaluation for external purposes also takes time, time that might be better spent addressing local questions and needs. Moreover, they felt that these questions and needs were precisely the ones in which the federal government should be interested but which were not reflected in the regulations.

Relative to the flexibility inherent in current federal requirements, directors mentioned that they had a hard time convincing their staff, resource specialists, and teachers to become involved in the design of the evaluation. Because of what they knew or believed about previous evaluations, personnel were hesitant about raising their own evaluation questions, reluctant to participate in the process of devising ways to answer them, and unwilling to believe that the evaluator would want to help them in the task of carrying out their day-to-day responsibilities. Thus, the current atmosphere of freedom from federal supervision and control -- far from stimulating staff members' interest in evaluation use -- seemed to arouse disbelief, anxiety, and hostility.

These observations seem to agree with the conclusions of our analyses of current federal requirements and their effects on local evaluations use.

Summary

In this paper we have drawn on our understanding of evaluation and its uses to examine current federal evaluation requirements.

It appears that, insofar as their language is concerned, Chapter 1 requirements are, as intended, potentially less burdensome than in the past. Chapter 2 regulations have even greater potential for increasing flexibility and decentralization. In both cases, the extent to which flexibility is increased depends, to some extent, on how states exercise their evaluation responsibilities. Some states may impose more rigid requirements than others, thus circumscribing the new freedom offered by federal regulations. Others may attempt to capitalize on their discretion but fail to stimulate local use. Therefore, tne role of the states should be considered in terms of the contribution it makes to promoting the local uses of evaluation.

These considerations aside, however, the language of current Chapter 1 and 2 requirements is unlikely to stimulate local programs to capitalize on their new found flexibility to promote local evaluation uses. The same criticism also applies to the language of the requirements accompanying bilingual, handicapped, and vocational education programs.

Now, the language of federal evaluation requirements should probably not begin to suggest ways of building local uses into program evaluations. Suggestions, as we have seen, have a tendency to be viewed as hard and fast requirements, and such a tendency runs counter both to the federal desire to reduce unnecessary burdens at the local level and to the need for local districts to influence the use factors that are important in their particular setting. On the other hand, federal regulatory language should be such that it does not create impediments, real or imagined, to local evaluation use. On this count, the language of current federal evaluation requirements needs to be examined.

Evaluation use does not happen without some careful nurturing, and the kind of nurturing that will stimulate local use, while it will likely vary in different program settings, can be affected by federal requirements. If the federal intention is to encourage local uses of evaluation information, or simply to permit local uses of evaluation, then the federal role in generating factors and factor interactions affecting such use needs to be recognized.

Figure 1

Factors Affecting Evaluation Use

I. HUMAN FACTORS

A. Evaluator Characteristics

1. commitment to use of evaluation results
2. view of desirability of user involvement in the evaluation
3. choice of role
4. rapport with users
5. understanding political considerations
6. credibility
7. background and identity

B. User Characteristics

1. identity
   a. range of users
   b. organizational positions
   c. professional experience levels
2. interest in the evaluation
   a. views about the program being evaluated
   b. program questions and concerns
   c. urgency of questions and concerns
   d. predisposition to evaluation/the evaluator
   e. perceived need for the evaluation
   f. expectations for the evaluation
   g. risks of the evaluation

3. commitment to use of evaluation results

4. professional styles

   a. administrative and organizational skills

   b. initiative

   c. openness to new ideas or change

5. information processing

II. CONTEXTUAL FACTORS

A. Preexisting Evaluation Bounds

1. written requirements for the evaluation

2. other contractual obligations for the evaluation

3. fiscal constraints

B. Organizational Characteristics

1. intraorganizational features

   a. role of central/district office

   b. interrelationship between unit and central/district administration

   c. institutional arrangements

   d. unit-level autonomy

   e. kinds of information, beyond evaluation, likely to be in use

   f. perceived institutional risk

2. extraorganizational features

   a. community climate

   b. community influence

   c. role of other agencies

C. Program Characteristics

1. age/maturity
2. innovativeness
3. overlap with other programs

III. EVALUATION FACTORS

A. Information Dialogue

1. amount of interaction between evaluator and users
2. quality of interaction between evaluator and users

B. Substance of Evaluation Information

1. information relevance for users
2. specific information selected for users

C. Evaluation Reporting

1. frequency of information provided
2. timing of information
3. format of presentations
4. statistical/narrative data
5. format of reports

D. Evaluation Procedures

1. methods used
   a. appropriateness
   b. rigor
2. dealing with mandated tasks
3. use of a general model

REFERENCES

Alkin, M.C., Daillak, R., & White, P. Using evaluations: Does evaluation make a difference? Beverly Hills, CA: Sage Publications, 1979.

Alkin, M.C., et al. Organizing for evaluation use: A handbook for administrators. Los Angeles: UCLA Center for the Study of Evaluation, 1983.

Burry, J. Evaluation in bilingual education. Evaluation Comment, 1979, 6,(1), 1-14.

Council of Chief State School Officers Consolidation Evaluation Task Force. Memorandum: Final report and recommendations. Council of Chief State School Officers, Washington, D.C.: 1982.

Daillak, R.H. A field study of evaluators at work. CSE Report No. 154. Los Angeles: UCLA Center for the Study of Evaluation, 1980.

Gray, P.J., Caulley, D.N., & Smith, N.L. A study in contrasts: Effects of the Education Consolidation and Improvement Act of 1981 on SEA and LEA evaluation. Report No. 79. Portland, Oregon: Northwest Regional Educational Laboratory, 1982.

Hastings, A.H. Snipping the strings: Local and state administrators discuss Chapter 2. Phi Delta Kappan, 1983, 65(3), 194-198.

Stecher, B.M., Alkin, M.C., & Flesher, G. Patterns of information use in school level decision making. CSE Report No. 160. Los Angeles: UCLA Center for the Study of Evaluation, University of California, December 1981.